



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/825,488	04/14/2004	Venkatesh Ganti	301560.01	8552
22971 7590 04/17/2009 MICROSOFT CORPORATION ONE MICROSOFT WAY REDMOND, WA 98052-6399				
			EXAMINER JOHNSON, JOHNESE T	
			ART UNIT 2166	PAPER NUMBER
			NOTIFICATION DATE 04/17/2009	DELIVERY MODE ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

roks@microsoft.com
ntovar@microsoft.com

Office Action Summary

Application No.

10/825,488

Applicant(s)

GANTI ET AL.

Examiner

Johnese Johnson

Art Unit

2166

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 05 January 2009.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-13, 15, 17-19, 21 and 25-34 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-13, 15, 17-19, 21, and 25-34 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☐ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date _____
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date _____
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: _____

DETAILED ACTION

Remarks

1. In response to the amendment filed on January 5, 2009, claims 1-21 and 25-34 are pending in this application.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

3. Claims 1-13, 15, 17-19, 21, and 25-34 are rejected under 35 U.S.C. 103(a) as being unpatentable over Borkar et al., "Automatic segmentation of text strings into structured records" and in view of Ando et al., "Mostly-Unsupervised Statistical Segmentation of Japanese Sequences".

As to claims 1 and 27, Borkar et al. disclose:

A process (see Abstract, pg. 1, line 1) and system (see Abstract, pg. 1, paragraph 2, line 1; wherein DATAMOLD is a system of interrelated components used to segment text) to evaluate an input string to segment said input string into component parts comprising:
means for providing a state transition model (see Abstract, pg. 1, paragraph 2, line 1
DATAMOLD) derived from training data from an existing collection of data records
that includes probabilities to segment input strings into component parts which

categorizes tokens in database attribute values of the data records into positions (see pg. 6, col. 2, lines 2-5 and 11-18), categorizes states for accepting classes of tokens into said positions, and adjusts said states and probabilities associated with said states within said positions to account for token placement in the input string, wherein training data corresponding to database attributes in (see pg. 7, section 2.5.1, lines 16-21); means for determining a most probable segmentation (see Abstract, pg. 1, paragraph 2, line 1 DATAMOLD) of the input string by comparing an order of tokens that make up the input string with a state transition model derived from the collection of data records (see pg. 3, section 1.3.1, col. 2, lines 9-11; wherein the inner HMMs corroborate each other's findings to pick the segmentation that is globally optimal).

means for segmenting the input string into one or more component parts according to the most probable segmentation (see page 4, col. 2, lines 6-9 and 37-38); and means for storing the one or more component parts in a data base (see abstract, line 7).

However, Borkar et al. do not explicitly disclose:

wherein the existing collection of data records does not comprise manually segmented training data.

Ando et al. disclose:

wherein the existing collection of data records does not comprise manually segmented training data (see abstract, lines 5-9 and page 2, lines 26-30).

It would have been obvious to have modified the teachings of Borkar et al. by the teachings of Ando et al. to provide a simple, efficient segmentation method thus avoiding the

costs of hand-segmenting (manually segmenting) training data (see Ando et al., page 2, lines 26-30).

As to claims 2 and 28, Borkar et al., as modified, disclose:

wherein the state transition model has probabilities for multiple states of said model and a most probable segmentation is determined based on a most probable token emission path through different states of the state transition model from a beginning state to an end state (see Borkar et al., pg. 4, col. 1, line 3; wherein the HMM has multiple states **and** col. 2, lines 6-9 –path having the highest probability).

As to claims 3 and 29, Borkar et al., as modified, disclose:

means for maintaining a collection of records, wherein the collection of data records is stored in a database relation and an order of attributes for the database relation as the most probable segmentation is determined (see Borkar et al., pg. 3, Fig. 1; wherein the structured record is determined and produced).

As to claims 4 and 30, Borkar et al., as modified, disclose:

wherein the input string is segmented into sub-components which correspond to attributes of the database relation (see Borkar et al., pg. 1, col. 2, section 1.1, lines 5-18).

As to claims 5 and 31, Borkar et al., as modified, disclose:

wherein the tokens that make up the input string are substrings of said input string (see Borkar et al., pg. 6, section 2.4, lines 2-4).

As to claims 6 and 32, Borkar et al., as modified, disclose:

wherein the input string is to be segmented into database attributes and wherein each attribute has a state transition model based on the contents of the database relation (see Borkar et al., pg. 4, Fig. 2; wherein each attribute has a transition in the model).

As to claims 7 and 33, Borkar et al., as modified, disclose:

wherein the state transition model has multiple states for a beginning, middle and trailing position within an input string (see Borkar et al., pg. 6, Fig. 6; wherein state “1” is the beginning, state “2” is the middle and state “3” is the trailing position).

As to claims 8 and 34, Borkar et al., as modified, disclose:

wherein the state transition model has probabilities for the states and a most probable segmentation is determined based on a most probable token emission [state] path through different states of the state transition model from a beginning state to an end state (see Borkar et al., pg. 6, Fig 6 and col. 2, paragraph 2, lines 1-4).

As to claim 9, Borkar et al., as modified, disclose:

wherein input attribute order for records to be segmented is known in advance of segmentation of an input string (see Borkar et al., Abstract, pg. 1, paragraph 2, lines 3-8).

As to claim 10, Borkar et al., as modified, disclose:

wherein an attribute order is learned from a batch of records that are inserted into the state transition model (see Borkar et al., Abstract, pg. 1, paragraph 2, lines 1-3).

As to claim 11, Borkar et al., as modified, disclose:

wherein the state transition model has at least some states corresponding to base tokens occurring in the reference relation (see Borkar et al., Abstract, pg. 1, paragraph 2, lines 1-8; wherein the training examples and dictionary provide the basis for acceptable and recognizable input and therefore some states would correspond to the same structure/ examples or base tokens).

As to claim 12, Borkar et al., as modified, disclose:

wherein the state transition model has class states corresponding to token patterns within said reference relation (see Borkar et al., pg. 3, col. 1, paragraph 3, lines 1-8).

As to claim 13, Borkar et al., as modified, disclose:

wherein the state transition model includes states that account for missing, misordered and inserted tokens within an attribute (see Borkar et al., pgs. 3-4, section 2; wherein data mold uses the example segmented records to output a model that when presented with any unseen text segments it into one or more of its constituent elements).

As to claim 15, Borkar et al., as modified, disclose:

A computer readable storage medium containing instructions that when executed cause a computer to perform the process of claim 1 (evaluat[ion] [of] an input string to segment said input string into component parts) (see Borkar et al., pg. 1, section 1.1, lines 5-6; wherein the tool is used during warehouse construction which implies that the program instructions are being read from a medium inserted in or stored on a machine).

As to claim 17, Borkar et al. disclose:

- a) a database management system to store records organized into relations
wherein data records within a relation are organized into a number of attributes
(see page 1, Abstract, line 7 – corporate database);
- b) a model building component that builds a number of attribute recognition
models derived from training data from an existing relation of data records,
wherein in one or more of said
attribute recognition models includes probabilities for segmenting input strings
into component parts which categorizes tokens in database attribute values of the
data records into positions, categorizes states for accepting classes of tokens into
said positions, and adjusts said states and probabilities associated with said states
within said positions to account for erroneous entries within an input string (see
page 1, Abstract, lines 13-14; wherein DATAMOLD comprises a model building
component because its built on HMM; and, (see pg. 7, section 2.5.1, lines 16-21
accounting for invalid paths); and
- c) a segmenting component that receives an input string and determines a most

probable record segmentation by evaluating transition probabilities of states within the attribute recognition models built by the model building component (see page 2, section 1.3, lines 1-3; wherein DATAMOLD comprises a segmenting component).

However, Borkar et al. do not explicitly disclose:

wherein training data corresponding to database attributes in the existing collection of data records does not comprise manually segmented training data.

Ando et al. disclose:

wherein the existing collection of data records does not comprise manually segmented training data (see abstract, lines 5-9 and page 2, lines 26-30).

It would have been obvious to have modified the teachings of Borkar et al. by the teachings of Ando et al. to provide a simple, efficient segmentation method thus avoiding the costs of hand-segmenting (manually segmenting) training data (see Ando et al., page 2, lines 26-30).

As to claim 18, Borkar et al., as modified, disclose:

wherein the segmenting component receives a batch of evaluation strings and determines an attribute order of strings in said batch and thereafter assumes the input string has tokens in the same attribute order as the evaluation strings (see Borkar et al., Abstract, pg. 1, paragraph 2, lines 3-8; wherein the training examples are the batch of strings that provide a basis for the structure of strings).

As to claim 19, Borkar et al., as modified, disclose:

wherein the segmenting component evaluates the tokens in an order in which they are contained in the input string and considers state transitions from multiple attribute recognition models to find a maximum probability for the state of a token to provide a maximum probability for each token in said input string (see Borkar et al., pg. 4, section 2.1; wherein the segmenting component considers transitions from the multiple attribute states to find the maximum probability).

As to claim 21, Borkar et al., as modified, disclose:

wherein the model building component defines a start and end state for each model and accommodates missing attributes by assigning a probability for a transition from the start to the end state (see Borkar et al., pg. 6, Fig. 6).

As to claim 25, Borkar et al. disclose:

A process of segmenting a string input record into a sequence of attributes for inclusion into a database table comprising: wherein determining a most probable segmentation of the input string comprises:

considering a first token in the input string and determining a maximum state

probability for said first token based on state transition models for multiple data table attributes (see pg. 4, section 2.1; wherein the segmenting component considers transitions from the multiple attribute states to find the maximum probability); and

considering in turn subsequent tokens in the input string and determining

maximum state probabilities for said subsequent tokens from a previous token state until all tokens are considered (see pg. 4, section 2.1; wherein the segmenting component considers transitions from the multiple attribute states to find the maximum probability); and

wherein segmenting the input string comprises segmenting the input string record by assigning the tokens of the string to attribute states of the state transition models corresponding to said maximum state probabilities (see pg. 4, Fig. 2, wherein the model displays attributes represented by states and section 2.1; wherein the segmenting component considers transitions from the multiple attribute states to find the maximum probability .

However, Borkar et al. do not explicitly disclose:

wherein the state transition models are derived from training data from an existing collection of data records that does not comprise manually segmented training data.

Ando et al. disclose:

wherein the state transition models are based on an existing collection of data records (sequences) that does not comprise manually segmented training data (see abstract, lines 5-9 and page 2, lines 26-30).

It would have been obvious to have modified the teachings of Borkar et al. by the teachings of Ando et al. to provide a simple, efficient segmentation method thus avoiding the costs of hand-segmenting (manually segmenting) training data (see Ando et al., page 2, lines 26-30).

As to claim 26, Borkar et al., as modified, disclose:

additionally comprising determining an attribute order for a batch of string input records and using the order to limit the possible state probabilities when evaluating tokens in an input string (see Borkar et al., Abstract, pg. 1, paragraph 2, lines 1-3; wherein the structure and order] is learned from the training examples).

8. Claims 14 and 20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Borkar et al.; “Automatic segmentation of text strings into structured records”, in view of Ando et al., “Mostly-Unsupervised Statistical Segmentation of Japanese Sequences”, and further in view of Reed (U.S. Pat. No. 5, 095, 432).

As to claim 14, Borkar et al. and Ando et al., do not explicitly disclose:

wherein the state transition model has a beginning, a middle and a trailing state topology and the process of accounting for misordered and inserted tokens is performed by copying states from one of said beginning, middle or trailing states into another of said beginning, middle or trailing states.

However, Reed discloses:

wherein the state transition model has a beginning, a middle and a trailing state topology and the process of accounting for misordered and inserted tokens is performed by copying states from one of said beginning, middle or trailing states into another of said beginning, middle or trailing states (see col. 5, lines 1).

It would have been obvious, at the time of the invention, having the teachings of

Borkar et al., Ando et al., and Reed before him/her, to combine the steps as disclosed by Borkar et al. and Ando et al. with the feature as disclosed by Reed to enable grammar developers to use the familiar PSG formalism to compile their grammars into RVG for more efficient execution (see Reed, col. 2, lines 54-57).

As to claim 20, Borkar et al. and Ando et al., do not explicitly disclose:

wherein the model building component assigns states for each attribute for a beginning, middle and trailing token position (see pg. 4, Fig. 2; wherein the states are assigned to each attribute and pg. 6, Fig. 6; wherein states are assigned for first (beginning state), second (middle state), third (trailing state))

However, Borkar et al. does not explicitly disclose:

wherein the model building component relaxes token acceptance by the model by copying states among said beginning, middle and trailing token positions.

Reed discloses:

wherein the model building component relaxes token acceptance by the model by copying states among said beginning, middle and trailing token positions (see col. 5, lines 1; wherein states in the transition model are copied).

It would have been obvious, at the time of the invention, having the teachings of Borkar et al., Ando et al., and Reed before him/her, to combine the steps as disclosed by Borkar et al. and Ando et al. with the feature as disclosed by Reed to enable grammar developers to use the familiar PSG formalism to compile their grammars into RVG for more efficient execution (see Reed, col. 2, lines 54-57).

4. Claim 16 is rejected under 35 U.S.C. 103(a) as being unpatentable over Borkar et al.; “Automatic segmentation of text strings into structured records” in view of Ando et al., “Mostly-Unsupervised Statistical Segmentation of Japanese Sequences”, and further in view of Fairweather (U.S. PG. Pub. No. 2006/0235811).

As to claim 16, Borkar et al. disclose:

providing a reference table of string records that are segmented into multiple substrings

corresponding to database attributes (see Abstract, p. 1, paragraph 2, lines 1-3);

categorizing states for accepting classes of tokens into said positions (see pg. 6, col. 2, lines 2-5);

breaking the input record into a sequence of tokens, and determining a most probable

segmentation of the input record by comparing the tokens of the input record with

state models derived for attributes from the reference table (see pg. 3, section 1.3.1, col.

2, lines 9-11; wherein the inner HMMs corroborate each other's findings to pick the segmentation that is globally optimal).

However, Borkar et al. does not explicitly disclose:

wherein the reference table of string records does not comprise manually segmented training data.

analyzing the substrings within an attribute to provide a state model that assumes a

beginning, a middle and a trailing token topology for said attribute said topology

including a null token for an empty attribute component;

Ando et al. disclose:

wherein the reference table of string records (sequences) does not comprise manually segmented training data. (see abstract, lines 5-9 and page 2, lines 26-30).

It would have been obvious to have modified the teachings of Borkar et al. by the teachings of Ando et al. to provide a simple, efficient segmentation method thus avoiding the costs of hand-segmenting (manually segmenting) training data (see Ando et al., page 2, lines 26-30).

However, Borkar et al. and Ando et al. does not explicitly disclose:

analyzing the substrings within an attribute to provide a state model that assumes a beginning, a middle and a trailing token topology for said attribute said topology including a null token for an empty attribute component

Fairweather discloses:

analyzing the substrings within database attribute values of string records for an attribute during a training phase to provide a state model that categorizes the substrings within database attribute values into positions based on a beginning, a middle and a trailing token topology for said attribute said topology including a null token for an empty attribute component (see Fairweather, paragraph [0406], lines 8-9; wherein a the null pointer is returned because the token is null);

It would have been obvious, at the time of the invention, having the teachings of Borkar et al., Ando et al., and Fairweather before him/her, to combine the steps as disclosed by Borkar et al. and Ando et al. with the feature as disclosed by Fairweather to provide a system in which the content of the data itself actually determines the order of execution of statements in the

mining language and automatically keeps track of the current state (see Fairweather, paragraph [0004], lines 7-10).

Response to Arguments

5. Applicant's arguments filed January 5, 2009 have been fully considered but they are not persuasive.

Applicants argument that “there is no teaching or suggestion in Borkar of a state transition model that categorizes tokens in database attribute values of the data records into positions and categorizes states for accepting classes of tokens into said positions” is accepted but is not deemed persuasive.

The examiner interprets this limitation to mean that the state transition model accepts certain attributes from each tuple in certain positions. Based on this interpretation, Borkar does this at page 6, col. 2, lines 2-5 wherein certain positions within the model accept numbers, words, and delimiters from the address tuples.

Applicants argument that “there is no teaching or suggestion in Borkar of a state transition model that categorizes states for accepting classes of tokens into said positions and adjusts said states and probabilities associated with said states within said positions to account for erroneous entries within an input string” is accepted but is not deemed persuasive.

The examiner interprets this limitation to mean that the state transition model accepts certain attributes from each tuple in certain positions and there are emission probabilities

associated with states that account for mistakes in the input string. Borkar does this at page 7, section 2.5.1, lines 16-21 wherein the model accounts for invalid paths.

Applicant continues to argue that “Ando fails to teach or suggest at least categorizing tokens in database attribute values of the data records into positions, categorizing states for accepting classes of tokens into said positions, and adjusting said states and probabilities associated with said states within said positions to account for erroneous token placement in the input string”. This argument is acknowledged but is not deemed persuasive.

As recited above, Borkar is cited for disclosing these limitations, not Ando.

Applicant also argues that, “Ando does not disclose training data corresponding to database attributes in an existing collection of data records that does not comprise manually segmented training data”. This argument is acknowledged but is not deemed persuasive.

The examiner interprets this limitation to mean that training data comprising attribute values from an existing table is not manually segmented. Based on this interpretation, Ando does this at abstract, lines 5-9 and page 2, lines 26-30 wherein the training data comprises long sequences of unsegmented data.

In response to applicant's argument that there is no suggestion to combine the references, the examiner recognizes that obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge

generally available to one of ordinary skill in the art. See *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988) and *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992). In this case, for claims 14 and 20:

It would have been obvious, at the time of the invention, having the teachings of Borkar et al., Ando et al., and Reed before him/her, to combine the steps as disclosed by Borkar et al. and Ando et al. with the feature as disclosed by Reed to enable grammar developers to use the familiar PSG formalism to compile their grammars into RVG for more efficient execution (see Reed, col. 2, lines 54-57).

As for claim 16:

It would have been obvious, at the time of the invention, having the teachings of Borkar et al., Ando et al., and Fairweather before him/her, to combine the steps as disclosed by Borkar et al. and Ando et al. with the feature as disclosed by Fairweather to provide a system in which the content of the data itself actually determines the order of execution of statements in the mining language and automatically keeps track of the current state (see Fairweather, paragraph [0004], lines 7-10).

Conclusion

6. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after

the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

7. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Johnese Johnson whose telephone number is 571-270-1097. The examiner can normally be reached on 4/5/9.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Hosain Alam can be reached on 571-272-3978. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/J. J./
Examiner, Art Unit 2166

Application/Control Number: 10/825,488

Page 19

Art Unit: 2166

April 14, 2009

JJ

/Hosain T Alam/

Supervisory Patent Examiner, Art Unit 2166